

УДК 621.372.061:391.266

О.І. Рибін, А.Д. Мельник

**РОЗПІЗНАВАННЯ ГОЛОСНИХ ЗВУКІВ “а”, “о”, “у”, “е” УКРАЇНСЬКОЇ МОВИ****Вступ**

Розпізнавання звуків людської мови є однією з важливих задач при створенні “інтелектуальних” комп’ютерів. Розв’язання такої задачі відрізняється від задачі розпізнавання окремих слів [1], динамічно введених підписів [2, 3] внаслідок схожості фонограм звуків “а”, “о”, “у”, “е”, що зумовлює їх взаємну заміну в живій розмовній мові і правилах правопису в процесі розвитку мови.

Різні за звучанням слова та динамічно введені підписи в наш час вже достатньо успішно розрізняються за їх еталонами, проте розпізнавання окремих (особливо голосних) звуків все ще лишається задачею на перспективу.

Методи класифікації сигналів можна умовно поділити на три великих класи [4, 5]: 1) детерміновані; 2) імовірнісні; 3) кореляційні.

Оскільки сигнал має імовірнісний характер, перший клас розглядати не будемо. Кореляційні методи (за самою назвою) споріднені з класом імовірнісних методів і виділені в окрему групу як такі, що є найменш громіздкими (при певних обмеженнях [5]) і тому найбільш поширеними.

Одним із найпростіших методів кореляційної класифікації є погоджена фільтрація [6]. Обмеження на класичну лінійну погоджену фільтрацію на основі великого статистичного матеріалу розглянуто в праці [7]. Більш гнучкою і такою, що краще відповідає вимогам розпізнавання різних за звучанням слів та динамічно введеного підпису, є нормальна [8, 9] та нормалізована [10–21] погоджена фільтрація.

У працях [15, 16, 19] запропоновано методи погодженої фільтрації на базі методів нормалізації стандартних перетворень тестовим сигналом і самі методи нормалізації (нормалізація “за кроком” [10–15]) тестового сигналу за класичними ортогональними перетвореннями (нормалізація “за рівнем” [16–20]) та метод побудови нормального перетворення [8, 9]. Нор-

малізація перетворення або тестового сигналу полягає в тому, що спектр трансформант (трансформанта — базисна функція ортогонального перетворення) нормалізованого [10, 12] або нормального [9] перетворення для тестового сигналу має єдину трансформанту. Тому відхилення досліджуваного сигналу від тестового проявляється в появі ненульових трансформант з більшими номерами, а чисельною оцінкою (детермінованою або імовірнісною) ступеня подібності є коефіцієнт трансформант [8, 14, 17] (аналог коефіцієнта гармонік для визначення спотворень гармоніки в перетворенні Фур’є [22]). При цьому чисельною оцінкою подібності досліджуваного сигналу тестовому є “гострота” (величина, протилежна коефіцієнту трансформант).

**Постановка задачі**

Метою статті є застосування методу обчислення умовної імовірності приналежності сигналу до наданого класу на базі ортогонального розкладу кореляційної матриці (метод Карунена—Лоева [4, 5]) для розпізнавання звуків людської мови.

**Методи і результати досліджень**

Спочатку розглянемо можливості нормалізованої [10–21] погодженої фільтрації для розпізнавання голосних звуків “а”, “о”, “у”, “е”.

Як видно з рис. 1, на якому наведено в збільшеному масштабі фонограми окремих звуків “а” (рис. 1, а), “о” (рис. 1, б), “у” (рис. 1, в), “е” (рис. 1, г), вони складаються з майже однакових сегментів для того самого звуку, дуже подібних між собою для різних звуків. Така подібність звуків вимагає вибору методу класифікації, чутливого до змін графоелементів окремих сегментів фонограм.

При ідентифікації звуку “а” усі методи нормальної (або нормалізованої) фільтрації дали негативні результати.

Так, в слові “казахстан” з трьома звуками “а” було надійно знайдено лише той звук, на який падав наголос (рис. 2).

Більш чутливим до змін графоелементів є метод обчислення умовної імовірності приналежності сигналу до наданого класу на базі ортогонального розкладу кореляційної матриці (метод Карунена—Лоева [4, 5]).

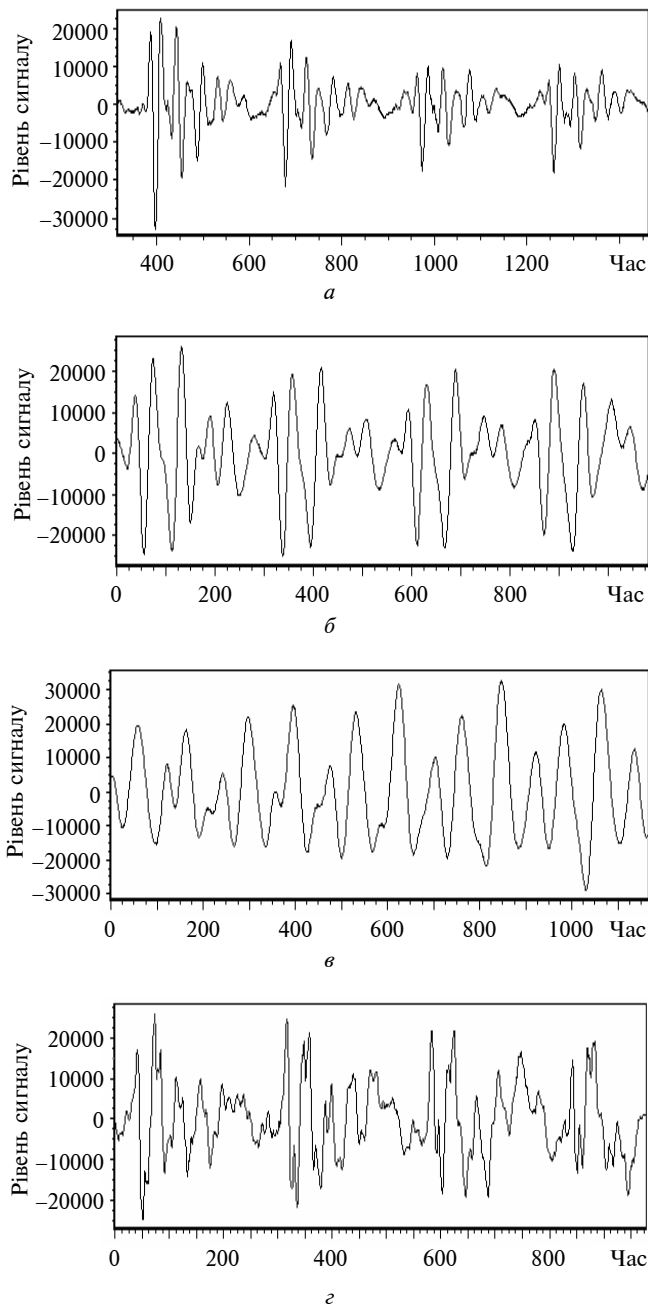


Рис. 1. Фонограми звуків у збільшеному масштабі: а – звук “а”; б – звук “о”; в – звук “у”; з – звук “е”

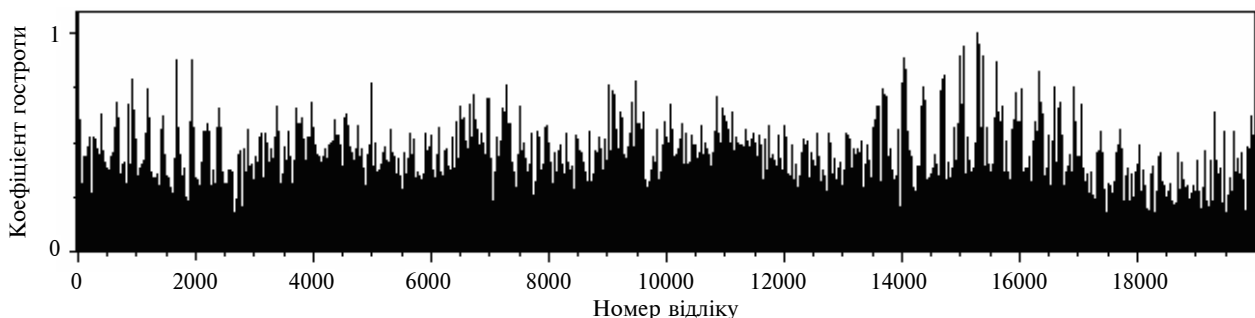


Рис. 2. Гострота досліджуваного сигналу у вікні, що посувалося вздовж фонограми “казахстан”

При цьому апостеріорна імовірність приналежності сигналу  $\bar{X}_0$   $n$ -го формату наданого класу (при умові, що це клас  $\omega_i$ ) обчислюється за формулою

$$P(\bar{X}_0/\omega_i) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(\overline{\text{Cor}}(\bar{X}))}} \times \exp\left(-\frac{1}{2}(\bar{X}_0 - \bar{M}_i)^T \overline{\text{Cor}}^{-1}(\bar{X})(\bar{X}_0 - \bar{M}_i)\right), \quad (1)$$

де  $T$  – знак транспонування;  $\bar{X}_0$  – вектор ознак досліджуваного сигналу (дискретні відліки фонограми невідомого звуку);  $\bar{M}_i$  – вектор середніх значень (математичне очікування) вектора  $\bar{X}$  ознак фонограм для класу  $\omega_i$  (наприклад, звуків “а”);  $\overline{\text{Cor}}(\bar{X})$  – кореляційна матриця сигналів класу  $\omega_i$ , тобто математичне сподівання добутку центрованих значень стовпця  $(\bar{X} - \bar{M}_i)$  і рядка  $(\bar{X} - \bar{M}_i)^T$  усіх навчальних сигналів класу  $\omega_i$ .

Як видно з (1), основну проблему при визначенні умовної імовірності  $P(\bar{X}_0/\omega_i)$  становить обернення матриці  $\overline{\text{Cor}}(\bar{X})$ , що разом з визначенням самої кореляційної матриці і є процедурою навчання класифікатора.

Як відомо [4, 5], кореляційна матриця сигналів певного класу (наприклад, фонограм звуків “а”, “о”, “у”, “е”) має певну структуру, при якій у головній діагоналі знаходяться найбільші числа – дисперсії випадкових відхилень від відповідних дискретних відліків математичних сподівань сигналу даного класу, а під та над діагоналлю – відповідні коефіцієнти кореляції, які зменшуються в міру віддалення від головної діагоналі. Таку матрицю  $\overline{\text{Cor}}$  (яка є

додатною напіввизначеною [23]) можна розкласти за власними значеннями  $\lambda_i$  і власними векторами  $\bar{\pi}_i$ :

$$\overline{\overline{\text{Cor}}} = \overline{\overline{\Pi}} \overline{\overline{\lambda}} \overline{\overline{\Pi}}^T, \quad (2)$$

де  $\overline{\overline{\Pi}}$ ,  $\overline{\overline{\Pi}}^T$  – матриці стовпців і рядків, утворених із власних векторів  $\bar{\pi}_i$ , відповідно;  $\overline{\overline{\lambda}}$  – діагональна матриця власних значень.

Під власними значеннями  $\lambda_i$  матриці  $\overline{\overline{\text{Cor}}}$  розуміють нулі визначника

$$|\overline{\overline{\text{Cor}}} - \lambda \overline{\overline{E}}|_{\lambda=\lambda_i} = 0, \quad (3)$$

де  $\overline{\overline{E}}$  – одинична матриця.

Під власними векторами розуміють розв'язки матричних рівнянь

$$[\overline{\overline{\text{Cor}}} - \overline{\overline{E}} \lambda_i] \bar{\pi}_i = \bar{0}, \quad (4)$$

де  $\bar{0}$  – стовпець нулів.

З виразу (2) можна зрозуміти джерела можливих похибок обернення матриці  $\overline{\overline{\text{Cor}}}$ , а також пояснити саму можливість такої процедури.

З (2) одержимо

$$\overline{\overline{\text{Cor}}}^{-1} = \overline{\overline{\Pi}} \overline{\overline{\lambda}}^{-1} \overline{\overline{\Pi}}^T. \quad (5)$$

Отже, оскільки ортонормовані матриці  $\overline{\overline{\Pi}}$ ,  $\overline{\overline{\Pi}}^T$  існують завжди, то обернення матриці  $\overline{\overline{\text{Cor}}}$  залежить від обернення матриці  $\overline{\overline{\lambda}}$ , яка є діагональною матрицею власних значень.

У випадку наявності зайвої (або корельованої) інформації у відліках  $\bar{X}$  одне або кілька з власних значень дорівнюватиме нулю (або “машинному” нулю). Таким чином, власні значення  $\lambda_i$  кореляційної матриці є маркером наявності або відсутності зайвої (збиткової) інформації.

Для обернення матриці  $\overline{\overline{\text{Cor}}}$  слід відкинути ці нульові (близькі до нуля) значення  $\lambda_i$  і відповідні їм власні вектори, а також привести одержані прямокутні матриці  $\overline{\overline{\Pi}}$ ,  $\overline{\overline{\Pi}}^T$  до квадратних відкиданням частини значень елементів власних векторів, номери яких у рядках (для  $\overline{\overline{\Pi}}^T$ ) і стовпцях  $\overline{\overline{\Pi}}$  відповідають номерам  $i$  відкинутих  $\lambda_i$ .

Але в цьому випадку порядок нової матриці  $\widetilde{\text{Cor}}$  буде меншим, ніж в (1). Оскільки зайві інформативні відліки для різних класів  $\omega_i$  сигналів виявляються різними за номерами, то процедура вибору еквідистантного кроку для цих всіх класів є досить громіздкою.

Якщо задачу обернення для всіх класів кореляційної матриці розв'язано, то оцінку приналежності досліджуваного сигналу до даного класу  $\omega_i$ ,  $i = 1, 2, \dots, M$ , проводять за дискримінантними числами  $D$  з (1):

$$D = (\bar{X}_0 - \bar{M}_i)^T \overline{\overline{\text{Cor}}}^{-1} (\bar{X}_0 - \bar{M}_i). \quad (6)$$

Якщо  $D \leq D_1$ , то  $\bar{X}_0 \in \omega_i$ ; якщо  $D > D_2$ ,  $D_2 > D_1$ , то  $\bar{X}_0 \notin \omega_i$ .

### Запропонований алгоритм класифікації

Згадані вище складності, пов'язані із створенням класифікаторів, ускладнюють їх навчання. Тому нами запропоновано метод класифікації, споріднений з методом за алгоритмом (1)–(6), але який не потребує обернення матриці  $\overline{\overline{\text{Cor}}}$ . Тоді всі досліджувані сигнали (нормовані по довжині до інтервалу  $[0, 1]$  та по миттєвих значеннях  $[-1, 1]$ ) будуть мати однакову кількість відліків і однакові формати матриць  $\overline{\overline{\text{Cor}}}$  для всіх класів.

Для пояснення запропонованого критерію класифікації розглянемо вираз (2), з якого випливає, що

$$\overline{\overline{\Pi}}^T \overline{\overline{\text{Cor}}} \overline{\overline{\Pi}} = \overline{\overline{\lambda}}. \quad (7)$$

Тепер введемо матрицю “миттєвої” кореляції

$$\widetilde{\text{Cor}} = (\bar{X}_0 - \bar{M}_i)(\bar{X}_0 - \bar{M}_i)^T, \quad (8)$$

усереднення за ансамблем якої для різних сигналів того самого класу  $\omega_i$  і дає матрицю  $\overline{\overline{\text{Cor}}}$ . Але в (8)  $\bar{X}_0$  – сигнал невідомого класу,  $\bar{M}_i$  – стовпець відліків математичного сподівання сигналів  $i$ -го класу. Усереднення не проводиться. Якщо тепер  $\bar{X}_0$  – відліки сигналу класу  $\omega_i$ , то з (7) отримаємо

$$\overline{\overline{\Pi}}^T \widetilde{\text{Cor}} \overline{\overline{\Pi}} = \tilde{\lambda}, \quad (9)$$

де  $\bar{\Pi}^T, \bar{\Pi}$  – матриці власних векторів для класу  $\omega_i$ ;  $\tilde{\lambda}$  – вже недиагональна матриця, але така, що

$$\|\bar{\lambda} - \tilde{\lambda}\| = \delta. \quad (10)$$

Якщо сигнал  $\bar{X}_0$  належить до класу  $\omega_i$ , то  $\delta < \delta_1$ , якщо ні, то  $\delta > \delta_2, \delta_2 > \delta_1$ . Тут  $\|\cdot\|$  означає норму (Евкліда, Хеммінга або Чебишова [4]). Як і в (6) для  $D_1, D_2$  порогові значення  $\delta_1, \delta_2$  визначають емпірично.

Результати досліджень за запропонованим алгоритмом

На базі алгоритму (8)–(10) було створено (по сегментах) окремі класифікатори, які в подальшому ці сегменти розпізнавали. В табл. 1

Таблиця 1. Результати застосування класифікаторів для розпізнавання звуків “а”, “о”, “у”, “е”

Номери сегментів сигналів	K <sub>λ</sub> · 10 <sup>3</sup> , “а”		K <sub>λ</sub> · 10 <sup>3</sup> , “о”		K <sub>λ</sub> · 10 <sup>3</sup> , “у”		K <sub>λ</sub> · 10 <sup>3</sup> , “е”	
	A1–A4	A5–A9	O1–O4	O5–O9	Y1–Y3	Y4–Y8	E1–E4	E5–E8
A1	<b>40.09</b>	129	309	525	748	609	337	647
A2	<b>10.77</b>	153	393	651	1026	732	503	868
A3	<b>15.97</b>	149	387	650	984	710	491	899
A4	<b>62.06</b>	128	395	677	1066	777	480	893
A5	20.06	<b>83</b>	341	629	1050	765	415	827
A6	127.03	<b>38</b>	261	554	1025	743	368	710
A7	216.75	<b>54</b>	278	515	968	681	317	595
A8	358.72	<b>42</b>	334	397	798	584	338	343
A9	411.58	<b>144</b>	508	514	858	463	498	512
Y1	1117	1071	621	1054	<b>35</b>	839	894	914
Y2	1072	1040	695	1227	<b>80</b>	844	927	1039
Y3	1353	1121	1128	1525	<b>139</b>	604	1036	1091
Y4	980	733	938	1004	611	<b>41</b>	746	693
Y5	838	648	848	972	537	<b>49</b>	715	719
Y6	726	616	858	1049	601	<b>60</b>	779	888
Y7	731	691	992	1249	847	<b>76</b>	960	1184
E1	357	236	254	399	624	588	<b>32</b>	392
E2	460	330	361	492	722	735	<b>28</b>	414
E3	614	477	403	612	954	900	<b>43</b>	442
E4	763	542	397	504	996	915	<b>121</b>	217
E5	738	466	466	384	1057	845	201	<b>76</b>
E6	935	588	512	386	928	846	325	<b>39</b>
E7	854	529	411	365	731	769	338	<b>50</b>
E8	1148	782	631	480	793	929	568	<b>77</b>
O1	513	381	<b>21</b>	564	607	880	370	461
O2	445	347	<b>21</b>	584	563	882	336	518
O3	397	303	<b>54</b>	367	676	868	329	533
O4	498	370	<b>95</b>	167	887	930	369	538
O5	590	445	261	<b>31</b>	1136	977	438	500
O6	627	457	341	<b>23</b>	1138	969	458	409
O7	624	430	318	<b>39</b>	1053	995	431	341
O8	720	503	450	<b>27</b>	1057	1034	463	317
O9	883	616	582	<b>90</b>	1147	1122	549	326

Таблиця 2. Усереднені результати застосування класифікаторів для розпізнавання звуків "а", "о", "у", "е"

Класифікатор	Результати застосування для звуків			
	"а"	"о"	"у"	"е"
А	<b>54.44</b>	454.78	784.44	533.11
О	478.22	<b>44.55</b>	655.44	366.33
У	890.0	959.71	<b>68.57</b>	905.87
Е	569.87	378.75	835.63	<b>58.25</b>

наведено результати випробувань класифікаторів для розпізнавання. При цьому кожен сигнал складався з семи–дев'яти сегментів. Так, для класифікатора звуку "а" – окремі (по сегментах) класифікатори А1–А4 та А5–А9, для сигналу "е" – Е1–Е4 та Е5–Е8 і т.д. По горизонталі в таблиці дано номери класифікаторів, по вертикалі – номери сегментів сигналів, які подавалися на ці класифікатори.

Результати з табл. 1 свідчать про те, що на "своїх" класифікаторах сегменти сигналів давали значно менші значення  $\delta$ , ніж на "чужих".

Підсумування результатів по сегментах і їх усереднення дає ще більш надійні результати класифікації (табл. 2).

### Висновки

У статті було продемонстровано невисоку чутливість методу фільтрації звуків "а", "о",

"у", "е" нормальним або нормалізованим погодженим фільтром. Альтернативним методом було запропоновано (на базі алгоритму перетворення Карунена–Лоева) спрощений алгоритм оцінки умовної імовірності приналежності сигналу до даного класу.

Одержані результати навчання запропонованих класифікаторів звуків "а", "о", "у", "е" за вибраним методом обробки показали високу чутливість запропонованого алгоритму класифікації.

У подальшому необхідно дослідити доцільність застосування створеного алгоритму оцінки умовної імовірності приналежності сигналу, який оброблюється, до даного класу сигналів для розпізнавання всіх голосних та приголосних звуків української мови.

А.И. Рыбин, А.Д. Мельник

РАСПОЗНАВАНИЕ ГЛАСНЫХ ЗВУКОВ "а", "о", "у", "э" УКРАИНСКОГО ЯЗЫКА

Предложены алгоритм классификации и классификаторы для идентификации звуковых сигналов. Высокая чувствительность классификаторов проиллюстрирована на примере распознавания звуков "а", "о", "у", "э".

O.I. Ribin, A.D. Melnyk

RECOGNITION OF VOWEL SOUNDS "a" "o" "u" "e" OF THE UKRAINIAN LANGUAGE

In this paper, we propose a classification algorithm for recognition of sound signals. Based on recognition of the vowel sounds "a", "o", "u", "e" of the Ukrainian language, we illustrate high sensitivity of the classification algorithm.

1. Єрмоленко Т.В. Применение вейвлет-преобразования для обработки и распознавания речевых сигналов // Искусственный интеллект. – 2002. – № 4. – С. 200–208.
2. Бойко І., Луценко С., Луцків А. Математичне моделювання та статистичні методи обробки динамічного підпису для задач аутентифікації особи в інформаційних системах // Електроніка та системи управління. – 2006. – № 2(8). – С. 27–37.
3. Луцків А.М. Імітаційне моделювання динамічно введеного підпису з метою тестування систем біометричної аутентифікації особи // Електроніка і зв'язь. – 2006. – № 5. – С. 44–48.
4. Абакумов В.Г., Рибін О.І., Сватош Й. Біомедичні сигнали. Генезис, обробка, моніторинг. – К.: Нора-прінт, 2001. – 516 с.
5. Продеус А.Н., Захарова Е.Н. Экспертные оценки в медицине. – К.: БЕК+, 1998. – 320 с.

6. *Финк Л.М.* Теория передачи дискретных сообщений. — М.: Сов. радио, 1970. — 728 с.
7. *Ян И.* Нелинейные согласованные фильтры для анализа различий // Изв. вузов. Радиоэлектроника. — 1999. — № 6. — С. 51–58.
8. *Рибін О.І., Ніжебецька Ю.Х.* Нормальне дискретне ортогональне перетворення // Вісн. НТУУ “КПІ”. Сер. Радіотехніка. Радіоапаратобудування. — 2008. — Вип. 37. — С. 8–15.
9. *Рибін О.І., Ніжебецька Ю.Х.* Алгоритм формування матричного оператора дискретного нормального перетворення // Там же. — С. 16–23.
10. *Рыбин А.И.* Нормализация дискретных ортогональных преобразований тестовым сигналом // Изв. вузов. Радиоэлектроника. — 2004. — № 7. — С. 39–46.
11. *Рыбин А.И., Григоренко Е.Г.* Алгоритм подстройки дискретного ортогонального преобразования под тестовый сигнал // Вісн. НТУУ “КПІ”. Сер. Приладобудування. — 2004. — № 27. — С. 122–128.
12. *Рибін О.І., Шарпан О.Б.* Діагностичні можливості процедури нормалізації ортогональних функцій при аналізі пульсограм // Вісн. ЖДТУ. Техн. науки. — 2004. — 1, № 4. — С. 144–149.
13. *Рибін О.І., Сакалош Т.В., Шарпан О.Б.* Аналіз пульсограм на базі процедури нормалізації ортогональних перетворень // Наукові вісті НТУУ “КПІ”. — 2005. — № 4. — С. 29–33.
14. *Рыбин А.И., Шарпан О.Б., Григоренко Е.Г., Сакалош Т.В.* Коэффициенты трансформант нормализованных ортогональных преобразований и диагностика пульсограм // Вісн. НТУУ “КПІ”. Сер. Приладобудування. — 2005. — Вип. 30. — С. 148–156.
15. *Рибін О.І., Данилевська В.Г.* Погоджена фільтрація на базі нормалізації ортогональних перетворень // Вісн. НТУУ “КПІ”. Сер. Радіотехніка. Радіоапаратобудування. — 2007. — Вип. 35. — С. 15–20.
16. *Рибін О.І., Мельник А.Д.* Погоджена фільтрація сигналів при зміні масштабу їх аргументів на базі нормалізованих вейвлет-функцій // Там же. — Вип. 34. — С. 18–24.
17. *Мельник А.Д., Рибін О.І.* Нормалізація тестового сигналу зі збереженням еквідистантного кроку дискретизації // Там же. — С. 24–29.
18. *Мельник А.Д., Рыбин А.И.* Нормализация эталонного сигнала с постоянным шагом дискретизации // Изв. вузов. Радиоэлектроника. — 2008. — № 1. — С. 71–75.
19. *Рыбин А.И., Мельник А.Д.* Согласованная нормализованная фильтрация сигналов // Там же. — № 2. — С. 77–80.
20. *Мельник А.Д., Рыбин А.И.* Согласованная вейвлет-фильтрация сигналов с измененным масштабом // Там же. — № 3. — С. 76–80.
21. *Рибін О.І., Мельник А.Д.* Алгоритм класифікації звукових сигналів // Вісн. НТУУ “КПІ”. Сер. Радіотехніка. Радіоапаратобудування. — 2008. — Вип. 37. — С. 19–27.
22. *Бессонов Л.А.* Теоретические основы электротехники. — М.: Наука, 2002. — 638 с.
23. *Гантмахер Ф.Р.* Теория матриц. — М.: Наука. — 1967. — 732 с.

Рекомендована Радою  
радіотехнічного факультету  
НТУУ “КПІ”

Надійшла до редакції  
24 жовтня 2008 року